Todd Drummond

The National Testing Initiative of the Kyrgyz Republic, Performance and Evaluation Plan

Award Number # 116-A-00-03-00014-00

SO 3.4.1 Democratic Values and Market Oriented Skills are Widely Integrated into the Curriculum

**USAID/CAR Regional Mission
Acquisition and Assistance Office
Department of State
7030, Almaty Place
Washington DC 20521-7030**

Grantee: American Councils for International Education: ACTR/ACCELS

July 29, 2003

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

1

The National Testing Initiative of the Kyrgyz Republic
Performance Monitoring and Evaluation Plan

Submitted July 29th, 2003

Overview

This performance monitoring and evaluation plan has been developed under contract 116-A-00-03-00014-00 and supports the objectives of the contract, described in the work plan as follows:

"The primary goal of this initiative is to promote the development of an independent testing organization (ITO) capable of providing professional assessment services in fair and transparent ways using modern educational practices and techniques."

Ultimately the ITO will be able to provide assessment services to all of the Central Asian Republics. Services can include, but are not limited to, high volume, high stakes testing, teacher assessments and primary school level assessments. It is expected that in the short term the ITO will execute the National Scholarship Test (NST) for university enrollment, thereby providing a transparent system of Government of Kyrgyzstan (GOK) scholarship distribution. A secondary objective of the project is to provide extensive guidance feedback to educators and donors based upon national scholarship test results. Findings based on these results can be utilized by teachers, administrators, policy makers and planners to improve educational policy as well as the teaching and learning processes. American Councils intends to provide guidance feedback to the educational community on the basis of NST results by the end of 2003.

This two part plan articulates expected outputs and results in the major sectors of the project. It is broken down into two parts. Part one set forth the targets (internal) that the project has set for itself and methods to measure attainment of those targets. These targets include training outputs, test administration and materials outputs, as well as financial goals of the new independent testing organization. Part two sets a plan for measuring *impact of the new testing and enrollment system on the education system as a whole*. First, goals are set for the level of public participation (transparency) in the new system. Second, data from the NST provides us the opportunity to measure access to higher

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

2

education from various demographic, linguistic and gender groups. Third, assessing fairness of the test can only be determined by assessing the validity of the testing instrument. A validation study will be conducted to see how well the selection test predicted (and thus selected) by comparing test scores with academic performance at a higher education institution. Finally, the project proposes a method for ascertaining the level of public support for the new testing and enrollment system.

Success in achieving targets for both parts one and two will be reported to USAID through the quarterly reporting system.

Quantitative Indicators of Project Team Performance

It is the scholarship test and processes before and after which provides the empirical test item, test, and personnel data in order to analyze project success. Numbers of people trained, tests or test brochures produced, study guides or other products produced, etc. can demonstrate a certain amount of effort on the part of the project team. They are concrete outputs. However, they are not the best indicators of the ultimate *results* of project team work. For example, the production and translation of testing brochures is an output. The effectiveness of provision of these testing brochures can only be assessed through the observation of behavioral results (better educated examinees is seen through fewer examinee mistakes for example). Creation of study guides is an output; measuring effectiveness of distribution through sales and income generation is a result. Fortunately, the project team in 2003 can use testing experience from 2002 as a baseline to measure improvement in results achieved from 2002 to 2003. American Councils conducted the test in 2002 on a subcontract from the CARANA Corporation. For example, in 2003 we will quantitatively measure improvements in time and resource efficiency in test administration from 2002. The same measurements can be made in 2004, utilizing 2002 and 2003 baseline data. We will ascertain increases or decreases in time needed to analyze test results, numbers of examinee mistakes due to instructional error, time administrators needed to administer the test, amount of resources expended versus materials actually utilized (tests ordered), etc. Measuring public opinion in support of the new system will also give us an idea of how the project has overcome corruption.

While the production of X amount of tests or test items is an output, we will measure statistically how each item performed. This analysis allows us to measure percentage of increase or decrease in efficiency of individual item developers. Comparisons of test performance can be made across working groups and subjects. The project team conducts such analysis on test items during "pre-testing" which gives us an indicator of how well test developers are progressing. Ability of a test to discriminate well between examinees and smooth "curve" distributions all indicate test quality and can be measured. So can appropriateness of difficulty level, feedback on typographical and content mistakes, etc. Therefore, while numbers of trained developers are outputs, the quality of their work is the result we are seeking.

As mentioned in the work plan, American Councils is not only measuring team performance in the science of testing, but also how organizational development of the

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

3

ITO is progressing. As in the science of testing, for organizational development we can measure both outputs and results, as well as note qualitative indicators of project success. The establishment of the ITO charter, mission statement, official registration, transfer of assets, meeting deadlines, and creation of a board of trustees are all measurable outputs for the first year of the project. Results can be measured by creation of income (increasing over time), diversification of services, procurement of outside funding and contracts, meeting training target goals, etc. Individual team member performance will be measured through the attainment of individual performance goals set by the team management. Achievement of special tax status is desirable but not necessarily a realistic, attainable, result due to the unpredictable political nature of this issue. USAID and other donors have worked in this area for over six years and are still struggling to develop comprehensible NGO tax law.

Qualitative Indicators of Project Team Performance

Public meetings, hearings before Parliament, seminars and trainings with educators, project lobbying through press and media campaigns are all in some form project outputs. While empirically measuring the success of these efforts is sometimes difficult, there are nonetheless ways to measure success. The project team envisions success in the public relations campaign to be measured in legislative support for the project through either Presidential and Ministerial decree or law- and to the extent that these various provisions incorporate AC and USAID project team recommendations. Specifically, the team is looking for transparency and openness in the university enrollment process as well as greater examinee freedom of choice in selecting major of choice (for government scholarships). Also, indicators of public support can be measured in public opinion surveys and in feedback of parents, teachers and students themselves. How well the public relations campaign is executed can also be measured in public response through registration for the NST in 2003 and 2004. High registration numbers and good geographic distribution will indicate that the message was received by the public. The project team will work with USAID to develop further appropriate mechanisms for measuring project performance in these areas.

Part One: Measuring Achievement of Project Targets for 2003-2005

American Councils, for the period of 2003-2005, has set a series of aggressive deliverables that will directly impact on the sustainability of the ITO in Kyrgyzstan. The following targets set for training and materials outputs, and milestones in developing the ITO below reflect realistic, necessary, and attainable goals. Targets are broken down into separate categories.

Training Output Targets for 2003-2005
The ITO needs to be able to access a pool of qualified, short term item developers for contract work. The ITO can not rely on only a few individuals. On the other hand, training too many potential item developers is neither realistic nor appropriate as the ITO does not have an interest in promoting the development of too many competing organizations. In terms of test administrators, the numbers below assume the same test

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

4

administration structure for the '04 that was used in '02 and '03. Registration may potentially be conducted by schools and trained registrars may not be required. The project team will advise as this becomes clearer.

Key for all following Charts	
Q1T	Quarter One Targets
Q1A	Quarter One Actual
Y1T	Year One Targets
Y1A	Year One Actual

Training Outputs: Deliverable Description		Year One: March 03 - March 04										
	Indicators	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y1T	Y1A
Deliverable #1	Main Test: Item Writers Trained (proficient)	30	0		0		15		15			**60**	
Deliverable #2	Main Test: Item Writer Trainers	0	0		0		6		0			**6**	
Deliverable #3	Subject Test: Item Writers Trained (proficient)	0	20		0		0		20			**40**	
Deliverable #4	Subject Test: Item Writer Trainers	0	0		0		4		0			**4**	
Deliverable #5	Test Administrators Trained	100	800		0		0		0			**900**	
Deliverable #6	Test Registrars Trained	0	150		0		0		0			**150**	

Training Outputs: Deliverable Description		Year Two: March 04 - March 05										
	Indicators	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y2T	Y2A
Deliverable #1	Main Test: Item Writers Trained (proficient)		0		15		15		0			**90**	
Deliverable #2	Main Test: Item Writer Trainers		0		0		4		0			**10**	
Deliverable #3	Subject Test: Item Writers Trained (proficient)		0		10		20		10			**80**	
Deliverable #4	Subject Test: Item Writer Trainers		0		0		8		0			**12**	
Deliverable #5	Test Administrators Trained		200		0		0		0			**1100***	
Deliverable #6	Test Registrars Trained		50		0		0		0			**200***	

*The number of administrators and registrars necessary for '04-05 will be determined only after the completion of a financial analysis of test administration '03. These numbers may be drastically reduced depending on the administration plan developed for '04-05. Also, in the event that nothing is altered, the number of trained administrators during the second year will be less than in '03 because ITO will utilize previously trained administrators. The '04-05 targets are final, cumulative targets over the life of the project.

Test Administration

The team will create three core examinee documents in order to conduct the NST-registration talons, answer sheets and score certificates. Registration talons are two part documents separated by a perforated line. This document serves as the primary transmitter of information from the examinee to the ITO and contains the most essential examinee information. There is also a place on the top and bottom halves for photographs. The ITO maintains one half while the other serves as an examinee entry ticket to testing as well as the ticket to collect individual score certificates. The development of quality answer sheets is essential. Answer sheets must be unambiguous,

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

5

neat, and accurate. The answer sheet serves as the primary transmitter of examinee testing effort to the ITO. In general, they are difficult to develop and require technical expertise. Finally, each examinee must receive a score certificate. This document is in two languages, contains (in 2004) five tear-off talons for university enrollment as well as an illustrated graph, depicting how the examinee performed relative to all other examinees. This document is collected and filed by universities upon enrollment. In 2004 this document was also color-coded to reflect "quota category" as requested by the Ministry of Education and Culture. The certificate also has a special "defense mechanism" which makes it difficult to forge.

In order to promote systems development for the ITO, core documents are needed to guide test administrators. The test administration handbook lays out all principles and procedures for test administrators, arms them with knowledge that protects them in difficult and unexpected situations, as well as provides them with guidance for troubleshooting and the safe handling of materials. It is the fundamental document for those in the field conducting testing. This handbook will always be somewhat of a work in progress but a core document that can be adapted is essential. Security protocols consist of 8-10 special documents that allow safe handling (check-in, check-out) of materials. Protocols for test administrators contain places for logging all essential information from test administration. Staff members refer to these protocols constantly during the "post-test" period. Careful documentation is the best way to demonstrate staff competence, prevent cheating, monitor irregularities and resolve difficult problems that arise. They also provide protection against the threat of lawsuits from the public.

Below are targets for testing materials outputs. "R" stands for Russian language, "K" for Kyrgyz language, and "U" for Uzbek language.

Test Administration Outputs: Deliverable Description (and language)		March 03 - March 04											
	Indicator	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y1T	Y1A
Deliverable #1	Answer Sheets (R,K), q=15,000 '02: 40,000 in '03	2	2		0		0		0			2*	
Deliverable #2	Tear Off Talon & Registration Sheets (R,K)	0	1		0		0		0			1	
Deliverable #3	Test Score Certificates Designed and Printed (R,K)	0	0		5		0		0			5	
Deliverable #4	Security Protocols (R)	1	1		0		0		0			1	
Deliverable #5	Test Administration Handbook (R), q= 800	0	1		0		0		0			1	

Test Administration Outputs: Deliverable Description (and language)		Year Two: March 04 - March 05											
	Indicator	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y2T	Y2A
Deliverable #1	Answer Sheets (R,K), q=15,000 '02: 40,000 in '03		2									2	
Deliverable #2	Tear Off Talon & Registration Sheets (R,K)			1								1	
Deliverable #3	Test Score Certificates Designed and Printed (R,K)			5								5	
Deliverable #4	Security Protocols (R)		1									1	
Deliverable #5	Test Administration Handbook (R), q= 800		1									1	
* as these documents need to be created every year , these totals													

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

6

are not cumulative over the life of the project.													

Testing Materials Output Targets for 2003-2005

In addition to the creation of tests of subject tests, the team will produce study guides which will be sold to generate income for the ITO. A shorter, less detailed overview of the test (brochure) will be produced that each examinee receives upon test registration. In 2002 and 2003, the project also produced a brochure on enrollment procedures that each examinee received upon completion of the test. In 2004, the Ministry will take responsibility for this task.

Test Outputs: Deliverable Description (and Language)		Year One: March 03 - March 04											
	Indicator	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y1T	Y1A
Deliverable #1	Core Tests (10 variations of R,K,U),	15	15		0		5		10			30*	
Deliverable #2	Chem/Bio. Subject Tests (10 each of R,K,U) +10	0	30		0		0		10			40*	
Deliverable #3	English/German Subject Tests, 12	0	8		0		0		4			12*	
Deliverable #4	Study Guides (R,K)**	0	1000**		0		3500		3500			8000	
Deliverable #5	Brochure on Test Rules (R,K,U)	0	1		0		0		0			1	
Deliverable #6	Brochure on Enrollment Procedures	1	1		0		0		0			1	

*Targets for Q1 are for 2003 test - tests developed in Q3 & Q4 are for the 2004 test. Test construction data is not cumulative over life of project.
**Targets indicate sales goals.

Test Outputs: Deliverable Description (and Language)		Year Two: March 04 - March 05											
	Indicator	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y2T	Y2A
Deliverable #1	Main Tests (10 variations of R,K,U), q= 42,000)		0		10		10		10			30*	
Deliverable #2	Chem/Bio. Subject Tests (10 each - R,K,U), 14,000		5		0		10		15			30*	
Deliverable #3	English/German Subject Tests, 8		4		0		2		2			8*	
Deliverable #4	Study Guides (R,K)**		5,000		3,000		2,000		2,000			12,000	
Deliverable #5	Brochure on Test Rules (R,K,U), q= 40,000		1		0		0		0			1	
Deliverable #6	Brochure on Enrollment Procedures, q= 40,000		0		0		0		0			0	

*Targets for Q1 are for 2004 test - tests developed in Q3&4 are for the 2005 test.
**Targets indicate sales goals.

ITO Development Outputs for 2003-05

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

7

Finally, the team has set establishment, training and development, and financial goals for the ITO team itself:

ITO Development Outputs: Deliverable Description		Year One: March 03 - March 04											
	Indicators	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y1T	Y1A
Deliverable #1	Mission Statement Completed	0	0		0		1		0			1	
Deliverable #2	Charter Completed	0	0		0		1		0			1	
Deliverable #3	ITO Registered and Functioning	0	0		0		1		0			1	
Deliverable #4	Board of Directors Installed	0	0		0		0		1			1	
Deliverable #5	Core Staff Training: Financial	0	0		0		0		0			0	
Deliverable #6	Core Staff Training: Management	0	0		0		0		0			0	
Deliverable #7	Core Staff Training: Fundraising	0	0		0		1		0			1	
ITO Development Outputs: Deliverable Description		Year Two: March 04-March 05											
	Indicators	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y2T	Y2A
Deliverable #5	Core Staff Training: Financial	0	0		0		1		0			1	
Deliverable #6	Core Staff Training: Management	0	0		0		1		0			1	

Revenue Targets: Deliverable Description	2003 Test		2004 Test								
	baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Q1T
1. Examination Fees	0	$46,195.65		.00	n/a	.		$120,487.50		$21,262.50
2. Teacher Training Methodology Courses	0	.		.00	n/a	.		.		.
3. Test Preparation Sales (15% purchase @ 3$/net)	0	.		.00	n/a	$1,620		$9,180		.
4. Teacher Training in Assessment	0	.		.00	n/a	.		.		.
Total Gross Revenue	0	$46,195.65		.00	n/a	$1,620		$129,667		$21,262.50
Total Net Revenue (gross revenue minus tax loss)	0	$32,337		.00	n/a	$1,482		$118,646		$19,455

Revenue Targets: Deliverable Description	2005 Test (after completion of contract)					
	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A
1. Examination Fees	.00		$21,150		$190,350	
2. Teacher Training Methodology Courses	$200		$400		$1,400	
3. Test Preparation Sales (15% purchase @ 3$/net	$324		$6,480		$3,996	
4. Teacher Training in Assessment	$60		$1,200		$740	
Total Gross Revenue	$584		$29,230		$195,486	
Total Net Revenue (gross revenue minus tax loss)	$534		$26,745		$170,109	

Progress on meeting the goal of developing ITO organizational capacity will be reported in quarterly reports submitted to USAID.

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

8

Part Two: Measuring Impact of the New Testing Regime on the Education System

Measuring Impact: Public Participation Targets (#1 below)

In accordance with USAID's SO 3.4.4 *Opportunities Exist for Public and Parental Participation in Education-related Decision Making,* the project team will work hard to ensure that the enrollment process developed by USAID, AC, and the Ministry is transparent. Fighting corruption in higher education has never been a highly complex technical matter. It is a matter of political will and the creation of sound systems. The team will propose both text for the Presidential and Ministerial decrees as well as advise MOK continually throughout the enrollment process. A significant accomplishment will be setting the precedent that scholarship enrollment is something that is not done behind closed doors. Up to the time of this contract, the project team has had tremendous success in getting the necessary conditions (NGO monitoring) institutionalized by MOK. This process began while AC was a subcontractor to CARANA on testing as early as the fall of 2002 through Parliamentary lobbying and preparation of the necessary Presidential and Ministerial decrees. The project team hopes to increasingly get more people involved as enrollment monitors throughout the life of the project. In 2003, the team hopes to involve at least 200 people from various NGO groups.

The team will also continue efforts to get MOK to seek out representation from various NGOs and the general public and promote dialogue on such important political questions as "scholarship quotas" and the selection mechanism. In 2003 a quota system will be used (Bishkek, small city, regional capitals, village, disability status). The implementation of this plan and future changes to it should only move forward with public participation and dialogue.

Public participation in the testing and enrollment process will be guaranteed through NGO monitoring of the university enrollment process. Targets for numbers of participants are set below as indicator #1. In 2002 there was no NGO monitoring of the distribution process. This was added in 2003. In 2002, only two members of the project team were on the MOK grant distribution committee. In 2003, over 200 NGO members will participate in monitoring the process at all universities around the country

Measuring Impact: Access to Higher Education by Gender (#2)

In accordance with USAID's SO 3.4.3 *Disadvantaged and At-risk Youth are Provided Adequate Access to Educational Opportunities,* the project team will measure access to higher education through the distribution of scholarship places by gender. Equal representation will be a by-product of both the Ministry's quota system and the fact that the test is merit-based. Outputs in support of this area include statistical reports on gender of both the examinee pool and those who become university entrants. As a larger proportion of women than men have higher education in Kyrgyzstan and girls outperformed boys on the NST in 2002 (59% of winners in the NST 2002), the issue of "girls education" is not necessarily as pressing an issue in Kyrgyzstan as it is in other

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

9

CAR countries. Girls were well represented in 2002 and are likely to be again in 2003. However, NST results will enable us to further monitor trends in this area closely.

As girls make up 52% of all students in higher education in the KR,[1] the project team seeks to ensure that females and males are equally represented in terms of scholarship slots. USAID and the project team need to consider new strategies to include more boys in the process, though the challenge is made more difficult by the fact that more than half of the students completing the 11th forms are girls. Statistics on gender by language, oblast, region, and test section will be presented and published in the report "The National Scholarship Test of the Kyrgyz Republic, 2003: A First Look." Data on gender will also be broken down between urban and rural areas.

Measuring Impact: Access to Higher Education by Demographics (#3)

Also in accordance with USAID's SO 3.4.3 *Disadvantaged and At-risk Youth are Provided Adequate Access to Educational Opportunitie*s, as a general rule, proportional representation by region (in terms of testing participation and scholarship receipt) is the goal of the project. Participation in testing will by measured by data taken from each oblast and region on numbers of participants involved in the test (indicator # 1). 86% of the school age population resides outside of the capital[2]. The project hopes to have at least at that numbers participating in the test as well as actually receiving scholarships. Obviously, however, ultimate success is not participation in testing but in receipt of scholarships.[3] Success will be measured by comparing numbers rural recipients in 2002 and 2003 against the previous year, 2001. 2001 numbers will serve as a baseline. We will not measure the total number of increase or decrease in places from these regions, but percentage of increase or decrease. This is because in 2001, there were more than 5085 scholarships offered.

Universities and the National Statistics Committee unfortunately do not maintain statistics on scholarship students by region of origin and therefore we will collect this data through regional departments of education. A sample of 20 regions (there are 63 total throughout the country) in regions of Russian, Kyrgyz, and Uzbek representation have been selected. Once collected, we will be able to measure increases or decreases from 2001. The baseline included here is 62% (village) but we also hope to see an 8-10% increase in rural representation from 2001 to 2002 which would indicate that test did impact scholarship distribution in favor of rural and impoverished regions during the first year. Quotas were also used in 2002 which should have helped rural students somewhat.

Getting public "buy in" for the testing initiative will also depend on the quality of detailed reports produced and made accessible to the general public and education community (#4). An overview of examinee results will be provided to USAID in the second quarterly report of the year. In the fall of 2003, a full informational report

---

[1] UNDP, Bishkek, 2002.
[2] While 86% of the population resides outside the capital, only 62% of these are defined as "rural" meaning students from villages, not small cities and regional centers. Thus, the concept of "rural" is a relative one.
[3] The effect of quotas on representation also needs to be considered.

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

10

complete with analysis will be produced for USAID summarizing the test results of the 2003 test. American Councils recommends making this report available both to the English and Russian speaking academic communities by publishing the report in the target numbers indicated below. This report will entail a twenty-five to thirty page overview of test results along with analysis for policy makers and educators.

Measuring Impact: Merit Based Selection to Higher Education

The team will collect data on academic performance for scholarship students from the years 2001, 2002, and 2003 in order to conduct a validation study of the NST. We ensured that the 2003 the Presidential Ukaz called for the collection of this data in order to allow us to conduct this work. This work is labor intensive as it requires university stakeholders to sort through a tremendous amount of data (in many cases not computerized) and will take time.

The team has selected a sample of 9 universities that are representative of the KR and will be collecting academic performance data in the winter of 2003. The overall justification for the use of the NST rests on the ability of the test instrument to predict the success of prospective students by providing a fair and unbiased evaluation of their knowledge and competency levels. Its purpose is to help educators identify among all students seeking entry those who are likely to excel. Predictive tests are evaluated by their predictive ability. A test validation study of the NST is essential in establishing it as a credible instrument that works. The team will look at the predictive validity of the test as a whole as well as be sections for key subject areas.

The role of American Councils will be to collect the data from all interested parties and American Councils' resources in Washington, DC, to write the validation study. In addition to the validity study, the team will continue to search for ways to demonstrate that high scoring students received scholarships. The team will also encourage universities to do their own studies and evaluations. For example in 2002, the Kyrgyz National University reported an improvement of 15% from the previous years. Their main indicators of success were grade point averages from the final first semester of study. They used their 2001 students as a baseline. Despite the potential for a host of intervening variables, differences in approach taken by each university to their own internal assessment systems, much can be learned by collecting and reporting this information. Universities are also eager to see the data from other universities. American Councils and the ITO can continue to act as the facilitator of this productive dialogue. This data can also be presented in a report format deemed acceptable to USAID.

Measuring Impact: Corruption in University Entrance

The steps taken to ensure openness and transparency of the NST and the enrollment process are well documented elsewhere (use of ID numbers, NGO monitoring, public disclosure of enrollment results, etc). The challenge however, is to measure to what extent these measures have actually impacted corruption. In general, this is a difficult accomplishment to measure. In order to do so the project will develop a questionnaire for

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

11

the key stakeholder group with the least amount of direct, personal stake in the enrollment system- school directors. Questions will be designed to elicit what directors think about how well the test discriminated between strong and weak students, as well as how well the system overcame corruption in the enrollment process. Directors will be asked to compare experience in 2002 and 2003 with 2001, the last year the old system was used. While the results will be somewhat anecdotal, it is widely accepted that corruption in higher education was endemic to the previous system. A strong vote of support in favor of the new systems over the old (we are shooting for 80% support as measured in terms of preferences for the current system over the older one) will be evidence that the new systems are tackling corruption and obstacles to fair access to higher education. Funding permitting, another survey will be conducted in 2005 to determine if support has increased after the 2003 test. The goal for this survey is 90% support for testing versus the previous, non-testing system.

Finally, in the winter of 2003-2004, American Councils will produce guidance feedback to the educational and donor communities in the form of an informational bulletin on the state of the teaching and learning processes as elicited from the results of NST 2003. Based on data elicited through the NST, American Councils has a wealth of information on areas for student, and thus teacher, performance. Guidance feedback goes the next step beyond information feedback (which is primarily statistical information on demographics) by providing informational feedback plus in depth analysis of examinee performance. For example, guidance feedback provides sample problems with their solutions. It articulates what and how the item is assessing in terms of examinee skills and knowledge. Then, simple classroom activities and methods are provided to help educators develop the skills or knowledge in question. High volume testing without guidance feedback is a wasted investment. The only way to improve the teaching and learning process is to learn from test results. Considering the importance of USAID's **Strategic Objective 3.4.1**, this is a crucial component of the project. The project team proposes to produce guidance feedback in several languages and to market this work, selling it at modest but realistic prices to the educational community. It is at once both a service for the ITO and a powerful instrument of educational reform. The project team hopes to complete its first bulletin by December 31st, 2003.

Measuring Impact of the New Testing Regime on the Education System

Impact of New Regime on Education System: Deliverable Description (and language)		Year One: March 03 - March 04											
	Indicator	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Stat.	Y1T	Y2A
Deliverable #1	Number of NGO monitors for Enrollment Process	2*			200							200	
Deliverable #2	Percentage of Girl Scholarship Winners	52%			50%							50%	
	Percentage of Rural Girl Scholarship Winners	N/A			50%							50%	
Deliverable #3	Percentage of non-Urban Scholarship Winners	62%			70%							70%	
Deliverable #4	Test Results Review (feedback for the general public: E,R,K)*	Not published					250					250**	
Deliverable # 5	Measuring Success Against Corruption (director survey)	N/A					80%					80% support	

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

12

Deliverable # 6	Guidance Feedback (report for educators: E,R,K)	N/A								500		500**	

* In 2002 the only observers of the enrollment process were project team representatives.
** Projected number of publications

Impact of New Regime on Education System: Deliverable Description (and language)		Year Two: March 04 - March 05											
	Indicators	Baseline	Q1T	Q1A	Q2T	Q2A	Q3T	Q3A	Q4T	Q4A	Status	Y2T	Y2A
Deliverable #1	Number of NGO monitors for Enrollment Process				250							250	
Deliverable #2	Percentage of Girl Scholarship Winners				50%							50%	
	Percentage of Rural Girl Scholarship Winners				50%							50%	
Deliverable #3	Percentage of non-Urban Scholarship Winners				70%							70%	
Deliverable #4	Test Results Review (feedback for the general public: E,R,K)*						500					500**	
Deliverable #5	Validation Study of Test Results		1	1								Acceptable Range by international standards	
Deliverable #6	Measuring Fight Against Corruption (director surveys)						90%					90%	

** Projected number of publications

Conclusion: Long Term Project Impact and Meeting USAID's Strategic Objectives

SO 3.4.1 Democratic Values and Market Oriented Skills are Widely Integrated into the Curriculum

There are other significant indicators of *long term* project success. The impact of merit based government scholarships will take many years to measure. However, there is no question that government investment in the most deserving students will have a long term economic benefit for the Republic. The guidance feedback we will provide teachers on areas for methodological improvement is also a longer term goal but its significance is clear. Thus, helping meet **USAID's SO 3.4.1** will probably be the most significant *long term* accomplishment of this project. The project team is already seeing the effect the test has had on teachers and students- demand for knowledge on how to be successful on the test, or "the wash-back effect" is now taking place. It will be important to cautiously harness that enthusiasm for knowledge about the test so that the wash-back effect does not result in teachers spending their entire lessons preparing students for the National Scholarship Test. "Teaching to the Test" can be a real danger, especially in conditions where access to higher education is competitive. That said, if the enthusiasm is harnessed, high stakes testing can serve as the motor of all education reform and the project team intend to monitor this situation carefully, working with other USAID and donor projects. A bit more on the potential impact of high stakes testing on the education system follows.

Harnessing the Power of High Stakes Testing to Promote Education Reform

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

13

Ten years since the creation of an independent Kyrgyz Republic, passivity, deference to authority, and intolerance of dissension are still inculcated to Kyrgyzstani youth in school and university classrooms. Teachers in Kyrgyzstan are utilizing the same methodological approaches they inherited/learned from their years of schooling in the Soviet era. While many of these techniques are good and have stood the test of time, many of them are more appropriate for "indoctrinating" students and creating dependence rather than providing them the skills and promoting the independent thinking that is necessary to participate in civil society. In fact, one could argue the case that these skills are highly discouraged in many instances. Authoritarian, teacher centered, top-down teaching methods need to be integrated with more student-centered, participatory approaches that promote the development of new skills and capabilities.

Techniques predominantly utilized in the classroom still include the very traditional teacher lecture, rote memorization of facts, retelling texts, dictations, translations and knowledge of rules. Questioning authority of teachers or canonical literature is not encouraged. What constitutes "knowledge" and even education is very "fact based," and "successful students" are those that have the capacity to retain standard cannons in the field. There are few opportunities to test and develop problem solving skills under real conditions. The problem is that knowing facts neither facilitates skill development nor enables one to address real world problems. Further, in the Kyrgyzstani classroom there is little in the way of student to student interaction, elicitation and consideration of student needs and feedback, consideration of unorthodox views, or development of skills needed in order to participate in civil society and resolve conflict. Students don't *participate, engage, problem solve, or debate, or even listen to their peers.* Instead, they "answer" memorized bodies of knowledge according to a predetermined plan with fixed, standard outcomes. Therefore, the NST and its focus on skills assessment is a unique and progressive intervention.

Some teachers in CAR have been receiving training in modern educational methods. Some of them have implemented new tools and methods and changed their attitudes towards their role in the classroom. However, due to the prevailing assessment systems, on the whole teachers have few incentives to utilize new methodologies learned from teacher training. This is because university entrance exams still require memorization and retention as the number one criteria for entry. As is most societies, parents hold teachers accountable for student performance on university entry exams. Now, for the first time, there is major incentive to completely alter classroom techniques to promote the necessary educational reform. In the words of the former Minister who initiated the testing project "students who can think will get scholarships." This was the rationale for developing a norm-based, aptitude test.

The project team is already seeing the visible impact of high stakes testing upon students and teachers. In all countries, there is an element of "teaching to the test." Teaching to *this* test however, requires a realignment of current practice. It provides incentives for teachers to learn techniques and methods and apply them in the classroom. We will likely continue to see major teacher initiative at the upper level as teachers are now striving to utilize "modern communicative and learner centered teaching methods (**SO 3.4.1**)" Only

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

14

the use of these methods will produce high test scores on the NST. The project team will be publishing guidance feedback for educators and policymakers which will help guide these classroom initiatives. It places tremendous responsibility on the project team as the wash-back effect can be potentially harmful. If the test continues to be conducted fairly, we could potentially see cases of massive abandonment of the traditional curricula in order that teachers spend their entire time preparing students for the scholarship test. We predict the demand for the type of teaching methods advocated by USAID in **SO 3.4.1** will increase dramatically in Kyrgyzstan over the coming years due to the National Scholarship Test. This will ultimately affect thousands of secondary schools at the upper levels.

The National Testing Initiative of the Kyrgyz Republic
American Councils for International Education
Performance Monitoring and Evaluation Plan

15